ELSEVIER

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Theoretical Computer Science 326 (2004) 117–135

Theoretical Computer Science

www.elsevier.com/locate/tcs

Random generation of $2\times 2\times\cdots\times 2\times J$ contingency tables

Tomomi Matsui[a,*], Yasuko Matsui[b], Yoko Ono[c]

[a] *Department of Mathematical Informatics, Graduate School of Information Science and Technology, University of Tokyo, Bunkyo-ku, Tokyo 113-0033, Japan.*

[b] *Department of Mathematical Sciences, School of Science, Tokai University, Kitakaname, Hiratsuka-shi, Kanagawa 259-1292, Japan.*

[c] *Department of Management Science, Faculty of Engineering, Tokyo University of Science, Shinjuku-ku, Tokyo 162-8601, Japan.*

Received 9 October 2002; received in revised form 9 April 2004; accepted 9 June 2004

Communicated by O. Watanabe

Abstract

We propose two Markov chains for sampling $(m+1)$-dimensional contingency tables indexed by $\{1,2\}^m \times \{1,2,\ldots,n\}$. Stationary distributions of our chains are the uniform distribution and a conditional multinomial distribution (which is equivalent to the hypergeometric distribution if $m=1$). Mixing times of our chains are bounded by $(\frac{1}{2})n(n-1)\ln(N/(2^m\varepsilon)) = (\frac{1}{2})n(n-1)\ln(dn/\varepsilon)$, where d is the average of the values in the cells and ε is a given error bound. We use the path coupling method for estimating the mixing times of our chains and showed that our chains mix rapidly.

Keywords: Contingency table; Markov chain Monte Carlo method; Rapidly mixing; Path coupling

1. Introduction

We propose two Markov chains for sampling $(m+1)$-dimensional contingency tables indexed by $\{1,2\}^m \times \{1,2,\ldots,n\}$. The first chain has the uniform distribution as a unique

* Corresponding author.

E-mail addresses: tomomi@misojiro.t.u-tokyo.ac.jp (T. Matsui), yasuko@ss.u-tokai.ac.jp (Y. Matsui), ono@ms.kagu.tus.ac.jp (Y. Ono)

URL: http://www.misojiro.t.u-tokyo.ac.jp/~tomomi/.

0304-3975/$ - see front matter
doi:10.1016/j.tcs.2004.06.014

stationary distribution. The stationary distribution of the second chain is a conditional multinomial distribution, which is the hypergeometric distribution for a 2-dimensional case. The mixing times of our chains are bounded by $(\frac{1}{2})n(n-1)\ln(dn/\varepsilon)$, where d is the average of the values in the cells and ε is a given error bound. We use the path coupling method [6,7] for estimating the mixing times of our chains.

Contingency tables are used in statistics to store data from sample surveys. Consider a scenario where N subjects are categorized into a table according to particular attributes. The data is often analyzed under the assumption that the attributes are independent; that is, the joint distribution is uniquely determined by the marginal probabilities. We often assume that each table is generated from the uniform distribution, or a conditional multinomial distribution (hypergeometric distribution) over the set of all the contingency tables (see [1,2][10,17], for example). One of the commonly used measures of independence is the χ^2 statistic [25]. A typical test of independence asks what fraction (that is, the sum of the probabilities) of the tables have a χ^2 value smaller than parameter t, as t varies. When the marginal totals are sufficiently large, we can apply the Pearson χ^2 test [25]. If the marginal totals include a small number, we need an exact inference for the contingency tables [17]. For the analysis of 2×2 contingency tables, an alternative to maximum likelihood estimation and χ^2 goodness-of-fit tests is Fisher's exact test for independence [18].

Fisher's exact test can be done by systematic enumeration of all the tables. However, when the number of tables is huge, exact enumeration is impractical. Mehta and Patel [24] proposed a network algorithm for the exact counting (not enumeration) of contingency tables. Aoki improved the network algorithm by using a path trimming technique [4]. However, the computational efforts and memory requirement of their algorithms are bounded by the table sum and are impractical when the table sum is large. For estimating the moments of the χ^2 statistic efficiently, a standard technique is the ordinary Monte Carlo method, if we have a method for sampling from the set of contingency tables. By using a rapidly mixing Markov chain with the desired stationary distribution, we can sample a contingency table after enough transitions of the Markov chain from an arbitrary initial state.

It is known that the problem of generating 3-dimensional contingency tables is intractable. More precisely, when we deal with 3-dimensional tables, the problem of checking the existence of at least one table that satisfies the given marginal totals is NP-complete [20]. Diaconis and Strumfels [13] proposed an algorithm for finding a Markov base for higher-dimensional contingency tables. Recently, Aoki and Takemura discussed Markov bases for some classes of 3-dimensional contingency tables [5,27]. In this paper, we deal with a special class of $(m+1)$-dimensional contingency tables, in which the cells are indexed by $\{1,2\}^m \times \{1,2,\ldots,n\}$. For this class, a natural Markov basis exists, which is a direct extension of the 2-dimensional case. This class of contingency tables arises in many practical situations [16,26]. There also exist some theoretical results on testing the independency of the attributes of $2\times 2\times K$ tables (see Agresti's survey paper [1], for example).

The problem of almost uniform sampling of contingency tables can be solved by using a Markov chain that converges to the uniform distribution. Diaconis and Saloff-Coste [12] discussed the rate of convergence of a natural Markov chain for 2-dimensional contingency tables. They showed that the ordinary chain mixes polynomial time in the table sum when the numbers of rows and columns are fixed. Dyer et al. [15] proposed a different Markov chain for counting the number of 2-dimensional contingency tables. In the case of sufficiently large marginal totals, their chain mixes polynomial time in the number of rows and columns. For 2-dimensional contingency tables with two rows, Hernek [19] showed that the mixing time of the ordinary Markov chain is bounded by a polynomial of the table sum and the number of columns. Hernek bounded the mixing time of the chain by using the coupling lemma [3]. Dyer and Greenhill [14] proposed a rapidly mixing Markov chain for two-row contingency tables. Their chain mixes polynomial time in the logarithm of the table sum and the number of columns. They analyzed the mixing rate of their chain by using the path coupling technique proposed by Bubley and Dyer [6,7]. Kannan et al. [23] gave a Markov chain with polynomial-time convergence for the 0-1 case with nearly equal marginal totals. In contrast, Chung et al. [8] proposed a Markov chain for contingency tables with large enough marginal sums; they also showed that their chain converges in pseudo-polynomial time. Recently, Cryan et al. [9] proposed a chain which is rapidly mixing when the number of rows (or columns) is constant.

We also consider the problem of generating contingency tables from a conditional multinomial distribution (hypergeometric distribution) over the set of all the contingency tables. In the 2-dimensional case, there exists a simple $\mathrm{O}(N)$ time perfect sampling method, where N is the table sum. However, the 3-dimensional case is computationally intractable in general, as described above. It is easy to see that the problem of sampling perfect matching in a given bipartite graph uniformly is a special case of the problem of generating $2 \times I \times I$ contingency tables from the conditional multinomial distribution, since in this case the conditional multinomial distribution becomes the uniform distribution. The existence of a polynomial time approximate uniform sampler for perfect matching was a long-standing open problem (see [21] for example). The problem was recently solved [22].

In Section 2, we introduce some notations and summarize the path coupling method. In Section 3, we describe our first chain whose stationary distribution is uniform. In Section 4, we discuss our second chain, whose stationary distribution is a conditional multinomial distribution.

2. Notations and definitions

We denote the set of integers (non-negative integers, positive integers) by Z (Z_+, Z_{++}), respectively. In this paper, we consider a set of $(m+1)$-dimensional contingency tables indexed by $\mathrm{B}^m \times J$ where $\mathrm{B} = \{1, 2\}$ and $J = \{1, 2, \ldots, n\}$. The all one vector in B^m is denoted by $\mathbf{1}$. Any index in J is called a *column index*. For any vector $\boldsymbol{x} \in \mathrm{Z}^{\mathrm{B}^m \times J}$, both $x(\boldsymbol{i}; j)$ and $x(i_1, i_2, \ldots, i_m; j)$ denote the element of $\boldsymbol{x}$ indexed by $\boldsymbol{i} = (i_1, i_2, \ldots, i_m) \in \mathrm{B}^m$ and $j \in J$. For any column index $j \in J$, $\boldsymbol{x}(j) \in \mathrm{Z}^{\mathrm{B}^m}$ denotes the subvector of $\boldsymbol{x} \in \mathrm{Z}^{\mathrm{B}^m \times J}$ consist of elements defined by indices in $\mathrm{B}^m \times \{j\}$. Given a vector of indices $\boldsymbol{i} \in \mathrm{B}^m$ and an index $l \in \{1, 2, \ldots, m\}$, $\boldsymbol{i}_{\bar{l}}$ denotes the vector of indices $(i_1, \ldots, i_{l-1}, i_{l+1}, \ldots, i_m) \in \mathrm{B}^{m-1}$ and we also denote the vector $\boldsymbol{i}$ by $(\boldsymbol{i}_{\bar{l}}, i_l)$ by changing the order of elements. For any vector $\boldsymbol{x} \in \mathrm{Z}^{\mathrm{B}^m \times J}$ and $l \in \{1, 2, \ldots, m\}$, $x(\boldsymbol{i}_{\bar{l}}, i_l; j)$ denotes the element $x(\boldsymbol{i}; j)$ where $\boldsymbol{i} = (\boldsymbol{i}_{\bar{l}}, i_l)$.

Next, we introduce the definition of contingency tables. Fig. 1 shows an example. Let $(\boldsymbol{r}^1, \boldsymbol{r}^2, \ldots, \boldsymbol{r}^m; \boldsymbol{c})$ be a sequence of non-negative integer vectors where $\boldsymbol{r}^l \in \mathrm{Z}_+^{\mathrm{B}^{m-1} \times J}$ for each $l \in \{1, 2, \ldots, m\}$ and $\boldsymbol{c} \in \mathrm{Z}_+^{\mathrm{B}^m}$. The element of $\boldsymbol{r}^l$ indexed by $(\boldsymbol{i}'; j) \in \mathrm{B}^{m-1} \times J$ is table

$x(1,1;1)$	$x(1,1;2)$	$x(1,1;3)$	$x(1,1;4)$	$x(1,1;5)$	$x(1,1;6)$
$x(1,2;1)$	$x(1,2;2)$	$x(1,2;3)$	$x(1,2;4)$	$x(1,2;5)$	$x(1,2;6)$

2	5	1	4	5	9
6	3	6	3	2	0

$x(2,1;1)$	$x(2,1;2)$	$x(2,1;3)$	$x(2,1;4)$	$x(2,1;5)$	$x(2,1;6)$
$x(2,2;1)$	$x(2,2;2)$	$x(2,2;3)$	$x(2,2;4)$	$x(2,2;5)$	$x(2,2;6)$

5	2	7	1	0	3
6	4	4	3	8	8

marginal sums

						$c(1,1)$
						$c(1,2)$
$r^2(1;1)$	$r^2(1;2)$	$r^2(1;3)$	$r^2(1;4)$	$r^2(1;5)$	$r^2(1;6)$	

						26
						20
8	8	7	7	7	9	

						$c(2,1)$
						$c(2,2)$
$r^2(2;1)$	$r^2(2;2)$	$r^2(2;3)$	$r^2(2;4)$	$r^2(2;5)$	$r^2(2;6)$	

						18
						33
11	6	11	4	8	11	

$r^1(1;1)$	$r^1(1;2)$	$r^1(1;3)$	$r^1(1;4)$	$r^1(1;5)$	$r^1(1;6)$
$r^1(2;1)$	$r^1(2;2)$	$r^1(2;3)$	$r^1(2;4)$	$r^1(2;5)$	$r^1(2;6)$

7	7	8	5	5	12
12	7	10	6	10	8

Fig. 1. An example of $\mathrm{B} \times \mathrm{B} \times J \quad (|J| = 6)$ table (denoted by $\boldsymbol{x}^*$).

denoted by $r^l(\boldsymbol{i}'; j)$. The set of contingency tables corresponding to $(\boldsymbol{r}^1, \boldsymbol{r}^2, \ldots, \boldsymbol{r}^m; \boldsymbol{c})$ is defined by

$$\mathcal{T} \stackrel{\text{def.}}{=} \left\{ \boldsymbol{x} \in \mathrm{Z}_+^{\mathrm{B}^m \times J} \;\middle|\; \begin{array}{ll} x(\boldsymbol{i}_{\bar{l}}, 1;\, j) + x(\boldsymbol{i}_{\bar{l}}, 2;\, j) & (\forall l \in \{1, 2, \ldots, m\}, \\ \quad = r^l(\boldsymbol{i}_{\bar{l}};\, j) & \qquad \forall \boldsymbol{i}_{\bar{l}} \in \mathrm{B}^{m-1},\ \forall j \in J), \\ \sum_{j \in J} x(\boldsymbol{i};\, j) = c(\boldsymbol{i}) & (\forall \boldsymbol{i} \in \mathrm{B}^m) \end{array} \right\}.$$

Each element in $\mathcal{T}$ is called a *table* for simplicity. In the following, the sum total of elements $(\sum_{\boldsymbol{i} \in \mathrm{B}^m} c(\boldsymbol{i}))$ is denoted by N. Clearly, for any table $\boldsymbol{x} \in \mathcal{T}$, the sum total of elements of $\boldsymbol{x}$ is equal to N.

In the rest of this section, we briefly review the path coupling technique proposed by Bubley and Dyer [7]. We use the technique in later sections to estimate the mixing times of our Markov chains. Here we deal with a Markov chain $\mathcal{M}$ with state space $\mathcal{T}$. Assume that $\mathcal{M}$ has a unique stationary distribution $\pi : \mathcal{T} \to [0, 1]$. For any probability distribution function π' on $\mathcal{T}$, define the *total variation distance* between π and π' to be

$$\mathrm{D_{TV}}(\pi, \pi') \stackrel{\text{def.}}{=} \max_{\mathcal{T}' \subseteq \mathcal{T}} \left| \sum_{\boldsymbol{x} \in \mathcal{T}'} \pi(\boldsymbol{x}) - \sum_{\boldsymbol{x} \in \mathcal{T}'} \pi'(\boldsymbol{x}) \right| = (1/2) \sum_{\boldsymbol{x} \in \mathcal{T}} |\pi(\boldsymbol{x}) - \pi'(\boldsymbol{x})|.$$

If the initial state of the chain $\mathcal{M}$ is $\boldsymbol{x}$, we denote the distribution of the chain at time t by $P_{\boldsymbol{x}}^t : \mathcal{T} \to [0, 1]$, i.e.,

$$P_{\boldsymbol{x}}^t(\boldsymbol{y}) \stackrel{\text{def.}}{=} \Pr[X_t = \boldsymbol{y} \mid X_0 = \boldsymbol{x}] \qquad (\forall \boldsymbol{y} \in \mathcal{T}).$$

The rate of convergence to stationary from the initial state $\boldsymbol{x}$ may be measured by

$$\tau_{\boldsymbol{x}}(\varepsilon) \stackrel{\text{def.}}{=} \min\{t \mid \mathrm{D_{TV}}(\pi, P_{\boldsymbol{x}}^{t'}) \leqslant \varepsilon \ \text{ for all } \ t' \geqslant t\},$$

where the error bound ε is a given positive constant. The *mixing time* $\tau(\varepsilon)$ of $\mathcal{M}$ is defined by

$$\tau(\varepsilon) \stackrel{\text{def.}}{=} \max_{\boldsymbol{x} \in \mathcal{T}} \tau_{\boldsymbol{x}}(\varepsilon),$$

which is independent of the initial state.

Next, we define a special Markov process with respect to $\mathcal{M}$ called joint process. A *joint process* of $\mathcal{M}$ is a Markov chain (X_t, Y_t) defined on $\mathcal{T} \times \mathcal{T}$ satisfying that each of (X_t) and (Y_t), considered marginally, is a faithful copy of the original Markov chain $\mathcal{M}$. More precisely, we require that

$$\begin{aligned} \Pr[X_{t+1} = \boldsymbol{x}' | (X_t, Y_t) = (\boldsymbol{x}, \boldsymbol{y})] = P_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}'), \\ \Pr[Y_{t+1} = \boldsymbol{y}' | (X_t, Y_t) = (\boldsymbol{x}, \boldsymbol{y})] = P_{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{y}') \end{aligned}$$

for all $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{x}', \boldsymbol{y}' \in \mathcal{T}$ where $P_{\mathcal{M}}(\boldsymbol{x}, \boldsymbol{x}')$ and $P_{\mathcal{M}}(\boldsymbol{y}, \boldsymbol{y}')$ denotes the transition probability from $\boldsymbol{x}$ to $\boldsymbol{x}'$ and from $\boldsymbol{y}$ to $\boldsymbol{y}'$ of the original Markov chain $\mathcal{M}$, respectively.

Path coupling lemma (Bubley and Dyer [7]). *Let G be a directed graph with vertex set $\mathcal{T}$ and arc set $A \subseteq \mathcal{T} \times \mathcal{T}$. We assume that G is strongly connected. Let us introduce a positive integer length for each arc. For any ordered pair of vertices $(\boldsymbol{x}, \boldsymbol{x}')$ of G, the distance from $\boldsymbol{x}$* to *$\boldsymbol{x}'$, denoted by $\ell(\boldsymbol{x}, \boldsymbol{x}')$, is the length of the shortest path from $\boldsymbol{x}$* to *$\boldsymbol{x}'$, where the length of a path is the sum of the lengths of arcs in the path. Suppose that there exists a joint process $(X, Y) \mapsto (X', Y')$ with respect to $\mathcal{M}$ satisfying that*

$$1 > \exists \beta > 0, \ \forall (X, Y) \in A, \ \mathrm{E}[\ell(X', Y')] \leqslant \beta \ell(X, Y).$$

Then the mixing time $\tau(\varepsilon)$ of the original Markov chain $\mathcal{M}$ satisfies $\tau(\varepsilon) \leqslant (1-\beta)^{-1} \ln(D/\varepsilon)$ where D denotes the diameter of G with respect to ℓ, i.e., the distance of a farthest (ordered) pair of vertices.

3. Markov chain for uniform distribution

First, we show a lemma which implies an irreducible Markov chain defined on the set of tables $\mathcal{T}$. We define the *parity function* $\mathrm{p} : \mathrm{Z} \to \{1, -1\}$ by

$$\mathrm{p}(x) = \begin{cases} 1 & (x \text{ is an even integer }), \\ -1 & (x \text{ is an odd integer }). \end{cases}$$

For any index $\boldsymbol{i} \in \mathrm{B}^m$, we denote $\mathrm{p}(i_1 + i_2 + \cdots + i_m)$ by $\mathrm{p}(\boldsymbol{i})$. The vector $\Delta \in \{1, -1\}^{\mathrm{B}^m}$ is defined by $\Delta(\boldsymbol{i}) \stackrel{\text{def.}}{=} \mathrm{p}(\boldsymbol{i})$ for each index $\boldsymbol{i} \in \mathrm{B}^m$. Given an ordered pair of distinct column

0	−1	0	1	0	0
0	1	0	−1	0	0
0	1	0	−1	0	0
0	−1	0	1	0	0

$\Delta[4,2]$

2	**5**	1	**4**	5	9
6	**3**	6	**3**	2	0
5	**2**	7	**1**	0	3
6	**4**	4	**3**	8	8

$\boldsymbol{x}^*$

2	**4**	1	**5**	5	9
6	**4**	6	**2**	2	0
5	**3**	7	**0**	0	3
6	**3**	4	**4**	8	8

$\boldsymbol{x}^* + \Delta[4,2]$

Fig. 2. The vector $\Delta[4, 2]$.

indices (j', j''), we define the vector $\Delta[j', j''] \in \mathrm{Z}^{\mathrm{B}^m \times J}$ by

$$\Delta[j', j''](j) \stackrel{\text{def.}}{=} \begin{cases} \mathbf{0} & (j \in J \setminus \{j', j''\}), \\ \Delta & (j = j'), \\ -\Delta & (j = j''). \end{cases}$$

For any table $\boldsymbol{x} \in \mathcal{T}$, we introduce the set of neighboring tables (Fig. 2).

$$\mathrm{N}^0(\boldsymbol{x}) \stackrel{\text{def.}}{=} \{\boldsymbol{x}' \in \mathcal{T} \mid \exists (j', j'') \in J \times J, \quad j' \neq j'', \quad \boldsymbol{x}' = \boldsymbol{x} + \Delta[j', j'']\}.$$

It is easy to see that if $\boldsymbol{x}' = \boldsymbol{x} + \Delta[j', j'']$, then $\boldsymbol{x} = \boldsymbol{x}' - \Delta[j', j''] = \boldsymbol{x}' + \Delta[j'', j']$, and so $\boldsymbol{x}' \in \mathrm{N}^0(\boldsymbol{x})$ implies $\boldsymbol{x} \in \mathrm{N}^0(\boldsymbol{x}')$. For any pair of vectors $\boldsymbol{x}, \boldsymbol{x}' \in \mathrm{Z}^{\mathrm{B}^m \times J}$, $||\boldsymbol{x} - \boldsymbol{x}'||_1$ denotes the distance $\sum_{(\boldsymbol{i}; j) \in \mathrm{B}^m \times J} |x(\boldsymbol{i}; j) - x'(\boldsymbol{i}; j)|$ between $\boldsymbol{x}$ and $\boldsymbol{x}'$.

Lemma 1. *Let G^0 be an undirected graph with vertex set $\mathcal{T}$ and for any pair of vertices $\{\boldsymbol{x}, \boldsymbol{x}'\}$, there exists an edge between $\boldsymbol{x}$ and $\boldsymbol{x}'$ if and only if $\boldsymbol{x}' \in \mathrm{N}^0(\boldsymbol{x})$. Then the graph G^0 is connected, i.e., for any pair of vertices $\{\boldsymbol{x}, \boldsymbol{x}'\}$ of G^0, there exists a path on G^0 between $\boldsymbol{x}$ and $\boldsymbol{x}'$. The diameter (the distance of farthest pair of vertices) is less than or equal to $N/2^{m+1}$.*

Proof. Assume on the contrary that G^0 is not connected. Let $\{\boldsymbol{x}, \boldsymbol{x}'\}$ be a pair of vertices which minimizes $||\boldsymbol{x} - \boldsymbol{x}'||_1$ subject to the condition that there does not exist any path between $\boldsymbol{x}$ and $\boldsymbol{x}'$. Without loss of generality, we can assume that $\exists j' \in J, \ x(\mathbf{2}; j') < x'(\mathbf{2}; j')$, where $\mathbf{2}$ is the all-two vector in B^m. It directly implies the following:
1. $x(\boldsymbol{i}; j') < x'(\boldsymbol{i}; j')$ for any $\boldsymbol{i} \in \mathrm{B}^m$ satisfying $\mathrm{p}(\boldsymbol{i}) = \mathrm{p}(\mathbf{2})$,
2. $x(\boldsymbol{i}; j') > x'(\boldsymbol{i}; j')$ for any $\boldsymbol{i} \in \mathrm{B}^m$ satisfying $\mathrm{p}(\boldsymbol{i}) \neq \mathrm{p}(\mathbf{2})$,
3. $|x(\boldsymbol{i}; j') - x'(\boldsymbol{i}; j')| = |x(\mathbf{2}; j') - x'(\mathbf{2}; j')|$ for any $\boldsymbol{i} \in \mathrm{B}^m$.

Since $\sum_{j \in J} x(\mathbf{2}; j) = \sum_{j \in J} x'(\mathbf{2}; j)$, there exists a column index j'' satisfying $x(\mathbf{2}; j'') > x'(\mathbf{2}, j'')$. Then we have the following properties:
1. $x(\boldsymbol{i}; j'') > x'(\boldsymbol{i}; j'')$ for any $\boldsymbol{i} \in \mathrm{B}^m$ satisfying $\mathrm{p}(\boldsymbol{i}) = \mathrm{p}(\mathbf{2})$,
2. $x(\boldsymbol{i}; j'') < x'(\boldsymbol{i}; j'')$ for any $\boldsymbol{i} \in \mathrm{B}^m$ satisfying $\mathrm{p}(\boldsymbol{i}) \neq \mathrm{p}(\mathbf{2})$,
3. $|x(\boldsymbol{i}; j'') - x'(\boldsymbol{i}; j'')| = |x(\mathbf{2}; j'') - x'(\mathbf{2}; j'')|$ for any $\boldsymbol{i} \in \mathrm{B}^m$.

The vector $\boldsymbol{x}'' = \boldsymbol{x} + \Delta[j', j'']$ is non-negative and so $\boldsymbol{x}'' \in \mathcal{T}$. Since $\boldsymbol{x}'' \in \mathrm{N}^0(\boldsymbol{x})$ and that the pair $\{\boldsymbol{x}, \boldsymbol{x}'\}$ is disconnected, there does not exist any path between $\boldsymbol{x}''$ and $\boldsymbol{x}'$. The inequality $||\boldsymbol{x} - \boldsymbol{x}'||_1 > ||\boldsymbol{x}'' - \boldsymbol{x}'||_1$ contradicts with the minimality of $||\boldsymbol{x} - \boldsymbol{x}'||_1$.

The table $\boldsymbol{x}''$ defined above satisfies that $||\boldsymbol{x} - \boldsymbol{x}''||_1 = 2^{m+1}$. The above procedure decreases the distance between a distinct pair of vertices and the decrement is 2^{m+1}. If we apply the procedure $\lfloor ||\boldsymbol{x} - \boldsymbol{x}'||_1 / 2^{m+1} \rfloor$ times, the distance of two vertices is less than 2^{m+1}.

1	2	3	4	5	6
2	5	**1**	4	5	9
6	3	**6**	3	2	0
5	2	**7**	1	0	3
6	4	**4**	3	8	8

$\boldsymbol{x}^*$

1	3	1	3	1	3	1	3
3	0	**2**	**1**	1	2	0	3
5	7	**6**	**6**	7	5	8	4
4	8	**5**	**7**	6	6	7	5
7	3	**6**	**4**	5	5	4	6

$\mathcal{N}(\boldsymbol{x}^*, \{1,3\})$

Fig. 3. Vectors in $\mathcal{N}(\boldsymbol{x}^*, \{1, 3\})$.

3	5	**0**	4	5	9
5	3	**7**	3	2	0
4	2	**8**	1	0	3
7	4	**3**	3	8	8

$\boldsymbol{x}^* - \Delta[3,1]$

2	5	**1**	4	5	9
6	3	**6**	3	2	0
5	2	**7**	1	0	3
6	4	**4**	3	8	8

$\boldsymbol{x}^*$

1	5	**2**	4	5	9
7	3	**5**	3	2	0
6	2	**6**	1	0	3
5	4	**5**	3	8	8

$\boldsymbol{x}^* + \Delta[3,1]$

0	5	**3**	4	5	9
8	3	**4**	3	2	0
7	2	**5**	1	0	3
4	4	**6**	3	8	8

$\boldsymbol{x}^* + 2\Delta[3,1]$

Fig. 4. The set of neighbors $\mathrm{N}^1(\boldsymbol{x}^*, \{1, 3\})$.

It implies that the obtained pair of vertices are identical. Thus the diameter of G^0 is less than or equal to $N/2^{m+1}$. $\square$

The above lemma indicates the existence of an irreducible Markov chain on $\mathcal{T}$ such that the transition probability of an ordered pair of tables $(\boldsymbol{x}, \boldsymbol{x}')$ is positive if and only if $\boldsymbol{x}$ and $\boldsymbol{x}'$ are adjacent on G^0. When $m = 1$, this chain is a special case of the Markov chain proposed by Diaconis and Saloff-Coste [12]. However, as discussed in Dyer and Greenhill [14], the mixing rate of Diaconis and Saloff-Coste's chain is low. In the following, we describe our chain, which is an extension of the chain discussed by Dyer and Greenhill [14] for contingency tables with two rows.

For any table $\boldsymbol{x} \in \mathcal{T}$ and any pair of distinct column indices $\{j', j''\}$, we define the following set of tables (Fig. 3):

$$\begin{aligned}\mathcal{N}(\boldsymbol{x}, \{j', j''\}) \stackrel{\text{def.}}{=} \{&(\boldsymbol{y}(j'), \boldsymbol{y}(j'')) \in \mathrm{Z}_+^{\mathrm{B}^m \times \{j', j''\}} \mid \\ &\exists\, \theta \in \mathrm{Z},\ \ (\boldsymbol{y}(j'), \boldsymbol{y}(j'')) = (\boldsymbol{x}(j'), \boldsymbol{x}(j'')) + \theta(\varDelta, -\varDelta) \geqslant \boldsymbol{0}\}.\end{aligned}$$

Markov chain $\mathcal{M}^1$: We introduce our chain $\mathcal{M}^1$ with state space $\mathcal{T}$. For any table $\boldsymbol{x} \in \mathcal{T}$ and any pair of distinct column indices $\{j', j''\}$, we define the following set of tables:

$$\begin{aligned}\mathrm{N}^1(\boldsymbol{x}, \{j', j''\}) \stackrel{\text{def.}}{=} \{&\boldsymbol{x}' \in \mathcal{T} \mid \boldsymbol{x}'(j) = \boldsymbol{x}(j)\ (\forall j \in J \setminus \{j', j''\}), \\ &(\boldsymbol{x}'(j'), \boldsymbol{x}'(j'')) \in \mathcal{N}(\boldsymbol{x}, \{j', j''\})\} \\ = \{&\boldsymbol{x}' \in \mathcal{T} \mid \exists \theta \in \mathrm{Z},\ \boldsymbol{x}' = \boldsymbol{x} + \theta\, \varDelta[j', j''] \geqslant \boldsymbol{0}\}.\end{aligned}$$

Our first Markov chain $\mathcal{M}^1$ with the state space $\mathcal{T}$ is defined by the following transition procedure. We denote the state of the chain $\mathcal{M}^1$ at time t by X_t and the element of X_t indexed by $(\boldsymbol{i}; j)$ is denoted by $X_t(\boldsymbol{i}; j)$. Then the state X_{t+1} is determined as follows. First, choose a pair of distinct column indices $\{j', j''\}$ randomly. Next, choose a table X_{t+1} from $\mathrm{N}^1(X_t, \{j', j''\})$ at random (Fig. 4).

We estimate the mixing time of our chain $\mathcal{M}^1$. According to the definition, it is clear that $\mathcal{M}^1$ is aperiodic and irreducible. The transition probability of $\mathcal{M}^1$ from $\boldsymbol{x}$ to $\boldsymbol{y}$, denoted by $P_{\mathcal{M}^1}(\boldsymbol{x}, \boldsymbol{y})$ is

$$P_{\mathcal{M}^1}(\boldsymbol{x}, \boldsymbol{y}) = \begin{cases} \left(\binom{n}{2} |\mathcal{N}(\boldsymbol{x}, \{j', j''\})|\right)^{-1} & (\text{ if } \boldsymbol{y} \in \mathrm{N}^1(\boldsymbol{x}, \{j', j''\})), \\ \sum_{j'<j''} \left(\binom{n}{2} |\mathcal{N}(\boldsymbol{x}, \{j', j''\})|\right)^{-1} & (\boldsymbol{x} = \boldsymbol{y}), \\ 0 & (\text{otherwise}). \end{cases}$$

Since $P_{\mathcal{M}^1}(\boldsymbol{x}, \boldsymbol{y}) = P_{\mathcal{M}^1}(\boldsymbol{y}, \boldsymbol{x})$, the stationary distribution of the chain is uniform.

First, we introduce a directed graph G^1 with the vertex set $\mathcal{T}$ and the arc set $A = \{(\boldsymbol{x}, \boldsymbol{x}') \mid \boldsymbol{x}' \in \mathrm{N}^0(\boldsymbol{x})\}$. We define that the length of every arc in A is equal to 1. The distance of any ordered pair of vertices $(\boldsymbol{x}, \boldsymbol{x}')$ on G^1 is denoted by $\ell(\boldsymbol{x}, \boldsymbol{x}')$. Next, we define a joint process $(X, Y) \mapsto (X', Y')$ with respect to $\mathcal{M}^1$. For any pair of tables $(X, Y) \in A$, we define the transition probability of our joint process from (X, Y) to (X', Y'). Without loss of generality, we can assume that $X(1) \neq Y(1)$, $X(2) \neq Y(2)$ and $X(j) = Y(j)$ for all $j \in J \setminus \{1, 2\}$. In the joint process, we choose a pair of distinct column indices (j', j'').

Case 1: When $\{j', j''\} \subseteq \{3, \dots, n\}$, it is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$ and so we choose a pair $(Z(j'), Z(j''))$ from $\mathcal{N}(X, \{j', j''\})$ at random. We set X' and Y' to the contingency table obtained from X and Y by replacing $(X(j'), X(j''))$ and $(Y(j'), Y(j''))$ by $(Z(j'), Z(j''))$, respectively. Then, it is clear that (X', Y') is also in A and so $\ell(X', Y') = 1$.

Case 2: Consider the case that $\{j', j''\} = \{1, 2\}$. It is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$. We construct X' and Y' by using the same manner of Case 1. Then, we have $X' = Y'$ and $\ell(X', Y') = 0$.

Case 3: Consider the case that $\{j', j''\} = \{1, 3\}$. Without loss of generality, we can assume that $Y = X + \Delta[2, 1]$. Fig. 5 shows an example. For any column index j^*, we define the vector $\Delta[j^*] \in \mathrm{Z}^{\mathrm{B}^m \times J}$ by

$$\Delta[j^*](j) \overset{\text{def.}}{=} \begin{cases} \mathbf{0} & (j \in J \setminus \{j^*\}), \\ \Delta & (j = j^*). \end{cases}$$

Clearly, $\Delta[j', j''] = \Delta[j'] - \Delta[j'']$. Thus, $Y = X + \Delta[2] - \Delta[1]$ and

$$\begin{aligned} Y + \theta\Delta[3, 1] &= X + \Delta[2, 1] + \theta\Delta[3, 1] = X + \Delta[2] - \Delta[1] + \theta\Delta[3] - \theta\Delta[1] \\ &= X - (\theta + 1)\Delta[1] + \Delta[2] + \theta\Delta[3]. \end{aligned}$$

Now assume that $X + \theta\Delta[3, 1] \in \mathcal{T}$ for any integer $\theta \in \{L, L + 1, \dots, U + 1\}$. In the following, we show that $Y + \theta\Delta[3, 1] \in \mathcal{T}$ for any integer $\theta \in \{L, L + 1, \dots, U\}$. We can show the non-negativity of each column of $Y + \theta\Delta[3, 1]$ as follows:

$$(Y + \theta\Delta[3, 1])(1) = X(1) - (\theta + 1)\Delta[1](1) = (X + (\theta + 1)\Delta[3, 1])(1) \geqslant \mathbf{0},$$

$$(Y + \theta\Delta[3, 1])(2) = X(2) + \Delta[2](2) = Y(2) \geqslant \mathbf{0},$$

$$(Y + \theta\Delta[3, 1])(3) = X(3) + \theta\Delta[3](3) = (X + \theta\Delta[3, 1])(3) \geqslant \mathbf{0},$$

$$(Y + \theta\Delta[3, 1])(j) = Y(j) \geqslant \mathbf{0} \quad (\forall j \in \{4, 5, \dots, m\}).$$

3	5	**0**	4	5	9
5	3	**7**	3	2	0

4	2	**8**	1	0	3
7	4	**3**	3	8	8

$X^1 = X - \Delta[3,1]$
$= Y^1 + \Delta[1,2]$

2	5	**1**	4	5	9
6	3	**6**	3	2	0

5	2	**7**	1	0	3
6	4	**4**	3	8	8

$X^2 = X$
$= Y^2 + \Delta[1,2]$

1	5	**2**	4	5	9
7	3	**5**	3	2	0

6	2	**6**	1	0	3
5	4	**5**	3	8	8

$X^3 = X + \Delta[3,1]$
$= Y^3 + \Delta[1,2]$

0	5	**3**	4	5	9
8	3	**4**	3	2	0

7	2	**5**	1	0	3
4	4	**6**	3	8	8

$X^4 = X + 2\Delta[3,1]$

2	6	**0**	4	5	9
6	2	**7**	3	2	0

5	1	**8**	1	0	3
6	5	**3**	3	8	8

$Y^1 = Y - \Delta[3,1]$
$= X^2 + \Delta[2,3]$

1	6	**1**	4	5	9
7	2	**6**	3	2	0

6	1	**7**	1	0	3
5	5	**4**	3	8	8

$Y^2 = Y = X + \Delta[2,1]$
$= X^3 + \Delta[2,3]$

0	6	**2**	4	5	9
8	2	**5**	3	2	0

7	1	**6**	1	0	3
4	5	**5**	3	8	8

$Y^3 = Y + \Delta[3,1]$
$= X^4 + \Delta[2,3]$

$\Pr[(X',Y') = (X^1,Y^1) | (j',j'') = (1,3)] = 3/12,\quad \Pr[(X',Y') = (X^2,Y^1) | (j',j'') = (1,3)] = 1/12,$
$\Pr[(X',Y') = (X^2,Y^2) | (j',j'') = (1,3)] = 2/12,\quad \Pr[(X',Y') = (X^3,Y^2) | (j',j'') = (1,3)] = 2/12,$
$\Pr[(X',Y') = (X^3,Y^3) | (j',j'') = (1,3)] = 1/12,\quad \Pr[(X',Y') = (X^4,Y^3) | (j',j'') = (1,3)] = 3/12.$

Fig. 5. An example of Case 3-1.

Thus, we have shown that $Y + \theta\Delta[3,1] \in \mathcal{T}$ for any integer $\theta \in \{L, L+1, \ldots, U\}$. This property implies that $|\mathcal{N}(X, \{j', j''\})| - 1 \leqslant |\mathcal{N}(Y, \{j', j''\})|$.

In a similar way, we can show that $|\mathcal{N}(Y, \{j', j''\})| \leqslant |\mathcal{N}(X, \{j', j''\})| + 1$. Thus, the difference of the sizes of $\mathcal{N}(X, \{j', j''\})$ and $\mathcal{N}(Y, \{j', j''\})$ is either 0 or 1.

Case 3-1: Consider the case that $|\mathcal{N}(X, \{j', j''\})| \neq |\mathcal{N}(Y, \{j', j''\})|$. Without loss of generality, we can assume that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})| + 1$. By arranging the order of elements in $\mathrm{N}^1(X, \{j', j''\}) = \{X^1, X^2, \ldots, X^{k+1}\}$ and $\mathrm{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \ldots, Y^k\}$, we can assume that $X^1(\mathbf{1};1) > X^2(\mathbf{1};1) > \cdots > X^{k+1}(\mathbf{1};1)$ and $Y^1(\mathbf{1};1) > Y^2(\mathbf{1};1) > \cdots > Y^k(\mathbf{1};1)$, where $\mathbf{1}$ denotes the all one vector indexed by B^m. Then we choose (X', Y') as follows:

$$(X', Y') = \begin{cases} (X^i, Y^i) & \text{with probability} \\ & \quad (k-i+1)/k(k+1) \qquad \text{for } i \in \{1,2,\ldots,k\}, \\ (X^{i+1}, Y^i) & \text{with probability } i/k(k+1) \qquad \text{for } i \in \{1,2,\ldots,k\}, \end{cases}$$

where the sum total of the probabilities is equal to $(1+2+\cdots+k)/k(k+1) + (k+\cdots+2+1)/k(k+1) = 1$ (see Fig. 5 for example). In the following, we consider the case that $Y = X + \Delta[2,1]$. (We can deal with the case that $Y = X + \Delta[1,2]$ in a similar way.) As shown above, there exists a pair of integers L and U such that

$$\mathcal{N}(X, \{j', j''\}) = \{X + \theta\Delta[3,1] \mid \theta \in \{L, L+1, \ldots, U, U+1\}\},$$

$$\mathcal{N}(Y, \{j', j''\}) = \{Y + \theta\Delta[3,1] \mid \theta \in \{L, L+1, \ldots, U\}\}$$

and $X^i = X + (L-1+i)\varDelta[3,1]$, $Y^i = Y + (L-1+i)\varDelta[3,1]$. Then it is easy to see that

$$\begin{aligned} X^i = X + (L-1+i)\varDelta[3,1] &= Y - \varDelta[2,1] + (L-1+i)\varDelta[3,1] \\ &= Y^i - \varDelta[2,1] = Y^i + \varDelta[1,2]. \end{aligned}$$

Thus, we have $(X^i, Y^i) \in A$. We can also show that

$$\begin{aligned} Y^i = X^i + \varDelta[2,1] &= X^{i+1} - \varDelta[3,1] + \varDelta[2,1] \\ &= X^{i+1} - \varDelta[3] + \varDelta[1] + \varDelta[2] - \varDelta[1] \\ &= X^{i+1} + \varDelta[2,3] \end{aligned}$$

and $(X^{i+1}, Y^i) \in A$. From the above, $\ell(X', Y') = 1$ holds.

Case 3-2: Consider the case that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})|$.

We denote $\mathrm{N}^1(X, \{j', j''\}) = \{X^1, X^2, \ldots, X^k\}$ and $\mathrm{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \ldots, Y^k\}$. By arranging the order of the elements, we assume that $X^1(\mathbf{1};1) > X^2(\mathbf{1};1) > \cdots > X^k(\mathbf{1};1)$ and $Y^1(\mathbf{1};1) > Y^2(\mathbf{1};1) > \cdots > Y^k(\mathbf{1};1)$, where $\mathbf{1}$ denotes the all one vector in B^m. Then we choose (X', Y') randomly from $\{(X^1, Y^1), (X^2, Y^2), \ldots, (X^k, Y^k)\}$. It is easy to see that $(X', Y') \in A$ and so $\ell(X', Y') = 1$.

From the above cases, we have

$$\mathrm{E}[\ell(X', Y')] = \left(1 - \binom{n}{2}^{-1}\right).$$

It implies the following result.

Theorem 1. *The Markov chain $\mathcal{M}^1$ has the mixing time $\tau_1(\varepsilon)$ satisfying that*

$$\tau_1(\varepsilon) \leqslant (1/2)n(n-1)\ln(dn/(2\varepsilon)),$$

where d is the average of the values in cells, i.e., $d = N/(2^m n)$.

Proof. The diameter of the graph G^1 is equal to that of G^0 and so less than or equal to $N/2^{m+1}$. Path coupling lemma induces the desired result. $\square$

4. Markov chain for conditional multinomial distribution

In this section, we consider a conditional multinomial distribution given marginal sum. The distribution function $\psi : \mathcal{T} \to [0,1]$ is defined by

$$\psi(X) \stackrel{\text{def.}}{=} (1/\rho(\mathcal{T})) \prod_{(\boldsymbol{i};j)\in \mathrm{B}^m \times J} (X(\boldsymbol{i};j)!)^{-1} \quad \text{where}$$

$$\rho(\mathcal{T}) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{y}\in\mathcal{T}} \prod_{(\boldsymbol{i};j)\in \mathrm{B}^m \times J} (\boldsymbol{y}(\boldsymbol{i};j)!)^{-1}.$$

Next, we describe a Markov chain which has a conditional multinomial distribution as a stationary distribution.

Markov chain $\mathcal{M}^2$: First, we introduce a distribution function $\psi(\boldsymbol{x}, \{j', j''\})$ defined on $\mathrm{N}^1(\boldsymbol{x}, \{j', j''\})$ as follows

$$\psi(\boldsymbol{x}, \{j', j''\}) : Y \mapsto (1/\rho(\boldsymbol{x}; \{j', j''\})) \prod_{(\boldsymbol{i};j)\in \mathrm{B}\times\{j',j''\}} (Y(\boldsymbol{i}; j)!)^{-1},$$

where

$$\rho(\boldsymbol{x}; \{j', j''\}) \stackrel{\text{def.}}{=} \sum_{\boldsymbol{y}\in \mathrm{N}^1(\boldsymbol{x},\{j',j''\})} \prod_{(\boldsymbol{i};j)\in \mathrm{B}\times\{j',j''\}} (\boldsymbol{y}(\boldsymbol{i}; j)!)^{-1}.$$

The Markov chain $\mathcal{M}^2$ with the state space $\mathcal{T}$ is defined by the following transition procedure. We denote the state of the chain $\mathcal{M}^2$ at time t by X_t. Then the state X_{t+1} is determined as follows. First, choose a pair of distinct column indices $\{j', j''\}$ randomly. Next, choose a table X_{t+1} from $\mathrm{N}^1(X_t, \{j', j''\})$ under the distribution function $\psi(X_t, \{j', j''\})$.

According to the definition, it is clear that $\mathcal{M}^2$ is aperiodic and irreducible. Since the detailed balance equations hold, the function ψ is a unique stationary distribution function of $\mathcal{M}^2$.

We define a joint process $(X, Y) \mapsto (X', Y')$ with respect to $\mathcal{M}^2$. Recall that A is the arc set of digraph G^1 defined in the previous section. We also use the distance function ℓ. For any pair of tables $(X, Y) \in A$, we define the transition probability of our joint process from (X, Y) to (X', Y'). In the following, we consider the case that $(X, Y) \in A$. Without loss of generality, we can assume that $X(1) \neq Y(1)$, $X(2) \neq Y(2)$ and $X(j) = Y(j)$ for all $j \in J \setminus \{1, 2\}$. In the joint process, we choose a pair of distinct column indices (j', j'').

Case 1: When $\{j', j''\} \subseteq \{3, \ldots, n\}$, it is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$ and so we choose a pair $(Z(j'), Z(j''))$ from $\mathcal{N}(X, \{j', j''\})$ under the distribution function $\psi(X, \{j', j''\})$. We set X' and Y' to the contingency table obtained from X and Y by replacing $(X(j'), X(j''))$ and $(Y(j'), Y(j''))$ by $(Z(j'), Z(j''))$, respectively. Then, it is clear that (X', Y') is also in A and so $\ell(X', Y') = 1$.

Case 2: Consider the case that $\{j', j''\} = \{1, 2\}$. It is clear that $\mathcal{N}(X, \{j', j''\}) = \mathcal{N}(Y, \{j', j''\})$. We construct X' and Y' by using the same manner of Case 1. Then, we have $X' = Y'$ and $\ell(X', Y') = 0$.

Case 3: Consider the case that $j' = 1$ and $j'' = 3$. Other cases are treated in a similar way.

Case 3-1: Consider the case that $|\mathcal{N}(X, \{j', j''\})| \neq |\mathcal{N}(Y, \{j', j''\})|$. We can assume that $|\mathcal{N}(X, \{j', j''\})| > |\mathcal{N}(Y, \{j', j''\})|$ without loss of generality. Since $(X, Y) \in A$, it is easy to show that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})| + 1$. By arranging the order of elements in $\mathrm{N}^1(X, \{j', j''\}) = \{X^1, X^2, \ldots, X^{k+1}\}$ and $\mathrm{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \ldots, Y^k\}$, we can assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \cdots > X^{k+1}(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \cdots > Y^k(\mathbf{1}; 1)$.

Exactly one of the following two cases holds:

(i) $(Y^1(1), Y^2(1), \ldots, Y^k(1)) = (X^1(1), X^2(1), \ldots, X^k(1))$
and
$(Y^1(3), Y^2(3), \ldots, Y^k(3)) = (X^2(3), X^3(3), \ldots, X^{k+1}(3))$,

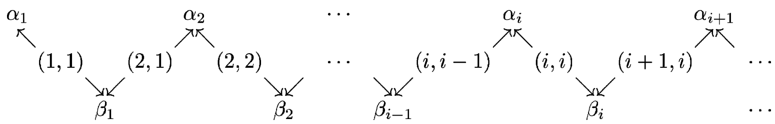

Fig. 6. Index set F.

(ii) $(Y^1(3), Y^2(3), \dots, Y^k(3)) = (X^1(3), X^2(3), \dots, X^k(3))$
and
$(Y^1(1), Y^2(1), \dots, Y^k(1)) = (X^2(1), X^3(1), \dots, X^{k+1}(1))$.

In the following, we consider Case (i). We can deal with Case (ii) in a similar way.

We choose (X', Y') as follows:

$$(X', Y') = \begin{cases} (X^1, Y^1) & \text{with probability } \psi(X, \{j', j''\})(X^1), \\ (X^i, Y^i) & \text{with probability } \sum_{i'=1}^{i} \psi(X, \{j', j''\})(X^{i'}) \\ & \quad - \sum_{i'=1}^{i-1} \psi(Y, \{j', j''\})(Y^{i'}) \quad \text{for } i \in \{1, 2, \dots, k\}, \\ (X^{i+1}, Y^i) & \text{with probability } \sum_{i'=1}^{i} \psi(Y, \{j', j''\})(Y^{i'}) \\ & \quad - \sum_{i'=1}^{i} \psi(X, \{j', j''\})(X^{i'}) \quad \text{for } i \in \{1, 2, \dots, k\}. \end{cases}$$

It is clear that the probabilities described above satisfy the equalities appearing in the definition of joint process. We need to show the non-negativity of the above probabilities. From the definition, $(X^i, Y^i), (X^i, Y^{i+1}) \in A$ for each $i \in \{1, 2, \dots, k\}$ and so $\ell(X', Y') = 1$. To show the non-negativity, we need the following lemma.

Lemma 2. *Let $(\alpha_1, \alpha_2, \dots, \alpha_{k+1})$ and $(\beta_1, \dots, \beta_k)$ be a pair of positive vectors satisfying that*

$$\frac{\alpha_1}{\beta_1} \leqslant \frac{\alpha_2}{\beta_2} \leqslant \cdots \leqslant \frac{\alpha_k}{\beta_k} \quad \textit{and} \quad \frac{\beta_1}{\alpha_2} \leqslant \frac{\beta_2}{\alpha_3} \leqslant \cdots \leqslant \frac{\beta_k}{\alpha_{k+1}}.$$

Let F be an index set defined by $F \stackrel{\text{def.}}{=} \{(1,1), (2,2), \dots, (k,k), (2,1), (3,2), \dots, (k+1,k)\}$ and $\gamma \in \mathbf{R}^F$ be a vector defined by

$$\begin{aligned} &\gamma(1,1) = \alpha_1/A, \\ &\gamma(i,i) = (\alpha_1 + \cdots + \alpha_i)/A - (\beta_1 + \cdots + \beta_{i-1})/B \qquad (i = 1, 2, \dots, k), \\ &\gamma(i+1,i) = (\beta_1 + \cdots + \beta_i)/B - (\alpha_1 + \cdots + \alpha_i)/A \quad (i = 1, \dots, k), \end{aligned}$$

where $A = \alpha_1 + \cdots + \alpha_{k+1}$ and $B = \beta_1 + \cdots + \beta_k$ (Fig. 6). Then the vector γ is non-negative.

Proof. (0) Clearly, $\gamma(1,1) \geqslant 0$.

(1) Let $i^* \stackrel{\text{def.}}{=} \max\{i \mid \beta_i/\alpha_{i+1} \leqslant B/A\}$. In the case that $\{i \mid \beta_i/\alpha_{i+1} \leqslant B/A\} = \emptyset$, we set $i^* \stackrel{\text{def.}}{=} -\infty$. For example, if $i^* \neq -\infty$, then we have the following inequalities

$$\frac{\beta_1}{\alpha_2} \leqslant \frac{\beta_2}{\alpha_3} \leqslant \cdots \leqslant \frac{\beta_{i^*}}{\alpha_{i^*+1}} \leqslant \frac{B}{A} < \frac{\beta_{i^*+1}}{\alpha_{i^*+2}} \leqslant \cdots \leqslant \frac{\beta_k}{\alpha_{k+1}}.$$

Now we show the non-negativity of $\gamma(i,i)$ for all i. For any index $i \in (-\infty, i^*] \cap \{1,2,\ldots,k\}$, $\gamma(i,i)$ satisfies

$$\begin{aligned}\gamma(i,i) &= (\alpha_1 + \cdots + \alpha_i)/A - (\beta_1 + \cdots + \beta_{i-1})/B \\ &= \alpha_1/A + (\alpha_2/A - \beta_1/B) + \cdots + (\alpha_i/A - \beta_{i-1}/B) \\ &\geqslant (\alpha_2/B)(B/A - \beta_1/\alpha_2) + \cdots + (\alpha_i/B)(B/A - \beta_{i-1}/\alpha_i) \geqslant 0.\end{aligned}$$

For any index $i \in [i^*+1, +\infty) \cap \{1,2,\ldots,k\}$, the following inequality holds:

$$\begin{aligned}\gamma(i,i) &= (\alpha_1 + \cdots + \alpha_i)/A - (\beta_1 + \cdots + \beta_{i-1})/B \\ &= (A - (\alpha_{i+1} + \cdots + \alpha_{k+1}))/A - (B - (\beta_i + \cdots + \beta_k))/B \\ &= (\beta_i/B - \alpha_{i+1}/A) + \cdots + (\beta_k/B - \alpha_{k+1}/A) \\ &\geqslant (\alpha_{i+1}/B)(\beta_i/\alpha_{i+1} - B/A) + \cdots + (\alpha_{k+1}/B)(\beta_k/\alpha_{k+1} - B/A) \geqslant 0.\end{aligned}$$

(2) Let i^* be the index defined by $i^* \stackrel{\text{def.}}{=} \max\{i \mid \alpha_i/\beta_i \leqslant A/B\}$. If $\{i \mid \alpha_i/\beta_i \leqslant A/B\} = \emptyset$, we define $i^* = -\infty$. For example, if $i^* \neq -\infty$ holds, we have the following inequalities:

$$\frac{\alpha_1}{\beta_1} \leqslant \frac{\alpha_2}{\beta_2} \leqslant \cdots \leqslant \frac{\alpha_{i^*}}{\beta_{i^*}} \leqslant \frac{A}{B} < \frac{\alpha_{i^*+1}}{\beta_{i^*+1}} \leqslant \cdots \leqslant \frac{\alpha_k}{\beta_k}.$$

Now we show that $\gamma(i+1,i) \geqslant 0$ for all i. For any index $i \in (-\infty, i^*] \cap \{1,2,\ldots,k\}$, $\gamma(i+1,i)$ satisfies

$$\begin{aligned}\gamma(i+1,i) &= (\beta_1 + \cdots + \beta_i)/B - (\alpha_1 + \cdots + \alpha_i)/A \\ &= (\beta_1/B - \alpha_1/A) + \cdots + (\beta_i/B - \alpha_i/A) \\ &= (\beta_1/A)(A/B - \alpha_1/\beta_1) + \cdots + (\beta_i/A)(A/B - \alpha_i/\beta_i) \geqslant 0.\end{aligned}$$

For any index $i \in [i^*+1, +\infty) \cap \{1,2,\ldots,k\}$, we can show the non-negativity of $\gamma(i+1,i)$ as follows:

$$\begin{aligned}\gamma(i+1,i) &= (\beta_1 + \cdots + \beta_i)/B - (\alpha_1 + \cdots + \alpha_i)/A \\ &= (B - (\beta_{i+1} + \cdots + \beta_k))/B - (A - (\alpha_{i+1} + \cdots + \alpha_{k+1}))/A \\ &= (\alpha_{i+1}/A - \beta_{i+1}/B) + \cdots + (\alpha_k/A - \beta_k/B) + \alpha_{k+1}/A \\ &\geqslant (\beta_{i+1}/A)(\alpha_{i+1}/\beta_{i+1} - A/B) + \cdots + (\beta_k/A)(\alpha_k/\beta_k - A/B) \geqslant 0.\end{aligned}$$

From the above, we have shown the non-negativity of the vector γ. $\square$

Let $\mathrm{B}^m_{\mathbf{1}} \stackrel{\text{def.}}{=} \{\boldsymbol{i} \in \mathrm{B}^m \mid \mathrm{p}(\boldsymbol{i}) = \mathrm{p}(\mathbf{1})\}$. For any index $\boldsymbol{i} \in \mathrm{B}^m_{\mathbf{1}}$, $(X^1(\boldsymbol{i};1), X^2(\boldsymbol{i};1), \ldots, X^{k+1}(\boldsymbol{i};1))$ is an arithmetic sequence of non-negative integers with the common difference -1. For each index $\boldsymbol{i} \in \mathrm{B}^m$, we denote the index obtained from $\boldsymbol{i}$ by flipping the mth element (i.e., the last element) by $\hat{\boldsymbol{i}}$. Then, for any index $\boldsymbol{i} \in \mathrm{B}^m_{\mathbf{1}}$, the sequence $(X^1(\hat{\boldsymbol{i}};1), X^2(\hat{\boldsymbol{i}};1), \ldots, X^{k+1}(\hat{\boldsymbol{i}};1))$ is an arithmetic sequence of non-negative integers with the common difference 1. It is also easy to show that for any index $\boldsymbol{i} \in \mathrm{B}^m_{\mathbf{1}}$, the sequences

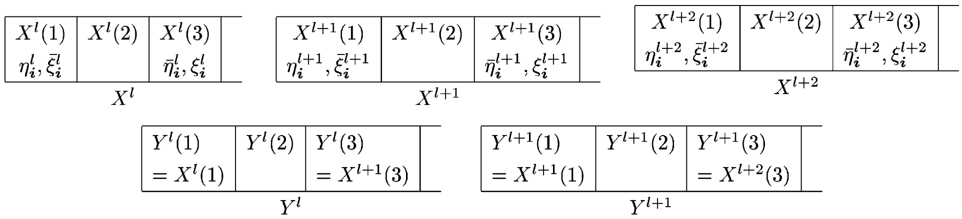

Fig. 7. Condition (i) of Lemma 3.

$(X^1(\boldsymbol{i};3), X^2(\boldsymbol{i};3), \ldots, X^{k+1}(\boldsymbol{i};3))$ and $(X^1(\hat{\boldsymbol{i}};3), X^2(\hat{\boldsymbol{i}};3), \ldots, X^{k+1}(\hat{\boldsymbol{i}};3))$ are arithmetic sequences of non-negative integers with the common differences 1 and -1, respectively. The above properties imply the following lemma.

Lemma 3. *Assume that the condition*

(i) $(Y^1(1), Y^2(1), \ldots, Y^k(1)) = (X^1(1), X^2(1), \ldots, X^k(1))$
and
$(Y^1(3), Y^2(3), \ldots, Y^k(3)) = (X^2(3), X^3(3), \ldots, X^{k+1}(3))$

is satisfied. Then the following inequalities hold:

$$\frac{\psi(X,\{j',j''\})(X^1)}{\psi(Y,\{j',j''\})(Y^1)} \leqslant \frac{\psi(X,\{j',j''\})(X^2)}{\psi(Y,\{j',j''\})(Y^2)} \leqslant \cdots \leqslant \frac{\psi(X,\{j',j''\})(X^k)}{\psi(Y,\{j',j''\})(Y^k)},$$

$$\frac{\psi(Y,\{j',j''\})(Y^1)}{\psi(X,\{j',j''\})(X^2)} \leqslant \frac{\psi(Y,\{j',j''\})(Y^2)}{\psi(X,\{j',j''\})(X^3)} \leqslant \cdots \leqslant \frac{\psi(Y,\{j',j''\})(Y^k)}{\psi(X,\{j',j''\})(X^{k+1})}.$$

Proof. We introduce some notations for simplicity. For any index $\boldsymbol{i} \in \mathrm{B}_{\mathbf{1}}^m$, we define

$$(\eta_{\boldsymbol{i}}^1, \eta_{\boldsymbol{i}}^2, \ldots, \eta_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\boldsymbol{i};1), X^2(\boldsymbol{i};1), \ldots, X^{k+1}(\boldsymbol{i};1)),$$

$$(\bar{\eta}_{\boldsymbol{i}}^1, \bar{\eta}_{\boldsymbol{i}}^2, \ldots, \bar{\eta}_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\hat{\boldsymbol{i}};3), X^2(\hat{\boldsymbol{i}};3), \ldots, X^{k+1}(\hat{\boldsymbol{i}};3)),$$

$$(\xi_{\boldsymbol{i}}^1, \xi_{\boldsymbol{i}}^2, \ldots, \xi_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\boldsymbol{i};3), X^2(\boldsymbol{i};3), \ldots, X^{k+1}(\boldsymbol{i};3)),$$

$$(\bar{\xi}_{\boldsymbol{i}}^1, \bar{\xi}_{\boldsymbol{i}}^2, \ldots, \bar{\xi}_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\hat{\boldsymbol{i}};1), X^2(\hat{\boldsymbol{i}};1), \ldots, X^{k+1}(\hat{\boldsymbol{i}};1)).$$

Then, both $(\eta_{\boldsymbol{i}}^1, \eta_{\boldsymbol{i}}^2, \ldots, \eta_{\boldsymbol{i}}^{k+1})$ and $(\bar{\eta}_{\boldsymbol{i}}^1, \bar{\eta}_{\boldsymbol{i}}^2, \ldots, \bar{\eta}_{\boldsymbol{i}}^{k+1})$ are arithmetic sequences of non-negative integers with common difference -1. Both of the sequences $(\xi_{\boldsymbol{i}}^1, \xi_{\boldsymbol{i}}^2, \ldots, \xi_{\boldsymbol{i}}^{k+1})$ and $(\bar{\xi}_{\boldsymbol{i}}^1, \bar{\xi}_{\boldsymbol{i}}^2, \ldots, \bar{\xi}_{\boldsymbol{i}}^{k+1})$ are arithmetic sequences of non-negative integers with common difference 1 (Fig. 7).

From the above, it is easy to see that

$$
\begin{aligned}
&\left(\frac{\psi(X,\{j',j''\})(X^l)}{\psi(Y,\{j',j''\})(Y^l)}\right)\left(\frac{\psi(X,\{j',j''\})(X^{l+1})}{\psi(Y,\{j',j''\})(Y^{l+1})}\right)^{-1}\\
&= \left(\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^l!\ \bar{\eta}_{\boldsymbol{i}}^l!\ \xi_{\boldsymbol{i}}^l!\ \bar{\xi}_{\boldsymbol{i}}^l!)^{-1}\right)\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^l!\ \bar{\eta}_{\boldsymbol{i}}^{l+1}!\ \xi_{\boldsymbol{i}}^{l+1}!\ \bar{\xi}_{\boldsymbol{i}}^l!)^{-1}\right)^{-1}\right)\\
&\quad\times\left(\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+1}!\ \bar{\eta}_{\boldsymbol{i}}^{l+1}!\ \xi_{\boldsymbol{i}}^{l+1}!\ \bar{\xi}_{\boldsymbol{i}}^{l+1}!)^{-1}\right)\right.\\
&\quad\times\left.\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+1}!\ \bar{\eta}_{\boldsymbol{i}}^{l+2}!\ \xi_{\boldsymbol{i}}^{l+2}!\ \bar{\xi}_{\boldsymbol{i}}^{l+1}!)^{-1}\right)^{-1}\right)^{-1}\\
&= \left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\bar{\eta}_{\boldsymbol{i}}^l)^{-1}\ \xi_{\boldsymbol{i}}^{l+1}\right)\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\bar{\eta}_{\boldsymbol{i}}^{l+1})^{-1}\ \xi_{\boldsymbol{i}}^{l+2}\right)^{-1}\\
&\quad\left[\begin{array}{l}\text{since }\ \bar{\eta}_{\boldsymbol{i}}^l > \bar{\eta}_{\boldsymbol{i}}^{l+1} > \bar{\eta}_{\boldsymbol{i}}^{l+2}\\ \qquad\quad\text{and } \xi_{\boldsymbol{i}}^l < \xi_{\boldsymbol{i}}^{l+1} < \xi_{\boldsymbol{i}}^{l+2}\end{array}\right]\\
&= \prod_{\boldsymbol{i}\in \mathrm{B}_1^m}\frac{\xi_{\boldsymbol{i}}^{l+1}\bar{\eta}_{\boldsymbol{i}}^{l+1}}{\bar{\eta}_{\boldsymbol{i}}^l\xi_{\boldsymbol{i}}^{l+2}} = \prod_{\boldsymbol{i}\in \mathrm{B}_1^m}\frac{\bar{\eta}_{\boldsymbol{i}}^{l+1}\xi_{\boldsymbol{i}}^{l+1}}{(\bar{\eta}_{\boldsymbol{i}}^{l+1}+1)(\xi_{\boldsymbol{i}}^{l+1}+1)} \leqslant 1.
\end{aligned}
$$

In a similar way, we can show that

$$
\begin{aligned}
&\left(\frac{\psi(Y,\{j',j''\})(Y^l)}{\psi(X,\{j',j''\})(X^{l+1})}\right)\left(\frac{\psi(Y,\{j',j''\})(Y^{l+1})}{\psi(X,\{j',j''\})(X^{l+2})}\right)^{-1}\\
&= \left(\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^l!\ \bar{\eta}_{\boldsymbol{i}}^{l+1}!\ \xi_{\boldsymbol{i}}^{l+1}!\ \bar{\xi}_{\boldsymbol{i}}^l!)^{-1}\right)\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+1}!\ \bar{\eta}_{\boldsymbol{i}}^{l+1}!\ \xi_{\boldsymbol{i}}^{l+1}!\ \bar{\xi}_{\boldsymbol{i}}^{l+1}!)^{-1}\right)^{-1}\right)\\
&\quad\times\left(\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+1}!\ \bar{\eta}_{\boldsymbol{i}}^{l+2}!\ \xi_{\boldsymbol{i}}^{l+2}!\ \bar{\xi}_{\boldsymbol{i}}^{l+1}!)^{-1}\right)\right.\\
&\quad\times\left.\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+2}!\ \bar{\eta}_{\boldsymbol{i}}^{l+2}!\ \xi_{\boldsymbol{i}}^{l+2}!\ \bar{\xi}_{\boldsymbol{i}}^{l+2}!)^{-1}\right)^{-1}\right)^{-1}\\
&= \left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^l)^{-1}\ \bar{\xi}_{\boldsymbol{i}}^{l+1}\right)\left(\prod_{\boldsymbol{i}\in \mathrm{B}_1^m}(\eta_{\boldsymbol{i}}^{l+1})^{-1}\ \bar{\xi}_{\boldsymbol{i}}^{l+2}\right)^{-1}\\
&= \prod_{\boldsymbol{i}\in \mathrm{B}_1^m}\frac{\bar{\xi}_{\boldsymbol{i}}^{l+1}\eta_{\boldsymbol{i}}^{l+1}}{\eta_{\boldsymbol{i}}^l\bar{\xi}_{\boldsymbol{i}}^{l+2}}\\
&= \prod_{\boldsymbol{i}\in \mathrm{B}_1^m}\frac{\eta_{\boldsymbol{i}}^{l+1}\bar{\xi}_{\boldsymbol{i}}^{l+1}}{(\eta_{\boldsymbol{i}}^{l+1}+1)(\bar{\xi}_{\boldsymbol{i}}^{l+1}+1)} \leqslant 1. \qquad \square
\end{aligned}
$$

If we set $\psi(X, \{j', j''\})(X^\ell) = \alpha_\ell$ and $\psi(Y, \{j', j''\})(Y^\ell) = \beta_\ell$, above lemmas directly imply the non-negativity of the transition probabilities of our joint process.

Case 3-2: Consider the case that $|\mathcal{N}(X, \{j', j''\})| = |\mathcal{N}(Y, \{j', j''\})|$. We denote $\mathrm{N}^1(X, \{j', j''\}) = \{X^1, X^2, \ldots, X^k\}$ and $\mathrm{N}^1(Y, \{j', j''\}) = \{Y^1, Y^2, \ldots, Y^k\}$. By arranging the order of the elements, we can assume that $X^1(\mathbf{1}; 1) > X^2(\mathbf{1}; 1) > \cdots > X^k(\mathbf{1}; 1)$ and $Y^1(\mathbf{1}; 1) > Y^2(\mathbf{1}; 1) > \cdots > Y^k(\mathbf{1}; 1)$. Exactly one of the following four cases holds:

$$\begin{aligned}
\text{Case (i)}:\ & (X^1(1), X^2(1), \ldots, X^k(1)) = (Y^1(1), Y^2(1), \ldots, Y^k(1)) \\
& \text{and} \\
& (X^2(3), X^3(3), \ldots, X^k(3)) = (Y^1(3), Y^2(3), \ldots, Y^{k-1}(3)), \\
\text{Case (ii)}:\ & (X^1(3), X^2(3), \ldots, X^k(3)) = (Y^1(3), Y^2(3), \ldots, Y^k(3)) \\
& \text{and} \\
& (X^2(1), X^3(1), \ldots, X^k(1)) = (Y^1(1), Y^2(1), \ldots, Y^{k-1}(1)), \\
\text{Case (iii)}:\ & (X^1(1), X^2(1), \ldots, X^k(1)) = (Y^1(1), Y^2(1), \ldots, Y^k(1)) \\
& \text{and} \\
& (X^1(3), X^2(3), \ldots, X^{k-1}(3)) = (Y^2(3), Y^3(3), \ldots, Y^k(3)), \\
\text{Case (iv)}:\ & (X^1(3), X^2(3), \ldots, X^k(3)) = (Y^1(3), Y^2(3), \ldots, Y^k(3)) \\
& \text{and} \\
& (X^1(1), X^2(1), \ldots, X^{k-1}(1)) = (Y^2(1), Y^3(1), \ldots, Y^k(1)).
\end{aligned}$$

In the following, we consider Case (i). We can deal with Cases (ii), (iii), (iv) in a similar way.

We choose (X', Y') as follows:

$$(X', Y') = \begin{cases}
(X^1, Y^1) & \text{with probability } \psi(X, \{j', j''\})(X^1), \\
(X^i, Y^i) & \text{with probability } \displaystyle\sum_{i'=1}^{i} \psi(X, \{j', j''\})(X^{i'}) \\
& \quad \displaystyle - \sum_{i'=1}^{i-1} \psi(Y, \{j', j''\})(Y^{i'}) \quad \text{for } i \in \{2, \ldots, k\}, \\
(X^{i+1}, Y^i) & \text{with probability } \displaystyle\sum_{i'=1}^{i} \psi(Y, \{j', j''\})(Y^{i'}) \\
& \quad \displaystyle - \sum_{i'=1}^{i} \psi(X, \{j', j''\})(X^{i'}) \quad \text{for } i \in \{1, 2, \ldots, k-1\},
\end{cases}$$

Clearly from the definition, $(X^i, Y^i), (X^{i+1}, Y^i) \in A$ for each $i \in \{1, 2, \ldots, k\}$ and so $\ell(X', Y') = 1$. We need to show the non-negativity of the probabilities defined above. To show the non-negativity, we need the following lemma.

Lemma 4. *Let $(\alpha_1, \ldots, \alpha_k)$ and $(\beta_1, \ldots, \beta_k)$ be a pair of positive sequences satisfying that*

$$\frac{\alpha_1}{\beta_1} \leqslant \frac{\alpha_2}{\beta_2} \leqslant \cdots \leqslant \frac{\alpha_k}{\beta_k} \text{ and } \frac{\beta_1}{\alpha_2} \leqslant \frac{\beta_2}{\alpha_3} \leqslant \cdots \leqslant \frac{\beta_{k-1}}{\alpha_k}.$$

Let F be an index set defined by $F \stackrel{\text{def.}}{=} \{(1,1),(2,2),\ldots,(k,k),(2,1),(3,2),\ldots,(k,k-1)\}$, *and* $\gamma \in \mathbf{R}^F$ *be a vector defined by*

$$\begin{aligned}\gamma(1,1) &= \alpha_1/A,\\ \gamma(i,i) &= (\alpha_1+\cdots+\alpha_i)/A - (\beta_1+\cdots+\beta_{i-1})/B \quad (i=2,3,\ldots,k),\\ \gamma(i+1,i) &= (\beta_1+\cdots+\beta_i)/B - (\alpha_1+\cdots+\alpha_i)/A \qquad (i=1,2,\ldots,k-1),\end{aligned}$$

where $A=\alpha_1+\cdots+\alpha_k$ *and* $B=\beta_1+\cdots+\beta_k$. *Then the vector* γ *is non-negative.*

Proof. We can show the lemma in a similar way with the proof of Lemma 2 by removing a_{k+1}. $\square$

Lemma 5. *Assume that the condition*

$$\begin{aligned}\mathit{Case}\ (\mathrm{i}):\ & (X^1(1),X^2(1),\ldots,X^k(1)) = (Y^1(1),Y^2(1),\ldots,Y^k(1))\\ & \mathit{and}\\ & (X^2(3),X^3(3),\ldots,X^k(3)) = (Y^1(3),Y^2(3),\ldots,Y^{k-1}(3))\end{aligned}$$

is satisfied. Then the following inequalities holds:

$$\frac{\psi(X,\{j',j''\})(X^1)}{\psi(Y,\{j',j''\})(Y^1)} \leqslant \frac{\psi(X,\{j',j''\})(X^2)}{\psi(Y,\{j',j''\})(Y^2)} \leqslant \cdots \leqslant \frac{\psi(X,\{j',j''\})(X^k)}{\psi(Y,\{j',j''\})(Y^k)},$$

$$\frac{\psi(Y,\{j',j''\})(Y^1)}{\psi(X,\{j',j''\})(X^2)} \leqslant \frac{\psi(Y,\{j',j''\})(Y^2)}{\psi(X,\{j',j''\})(X^3)} \leqslant \cdots \leqslant \frac{\psi(Y,\{j',j''\})(Y^{k-1})}{\psi(X,\{j',j''\})(X^k)}.$$

Proof. We introduce some notations for simplicity. For any index $\boldsymbol{i} \in \mathrm{B}_{\mathbf{1}}^m$, we define

$$(\eta_{\boldsymbol{i}}^1,\eta_{\boldsymbol{i}}^2,\ldots,\eta_{\boldsymbol{i}}^k) \stackrel{\text{def.}}{=} (X^1(\boldsymbol{i};1),X^2(\boldsymbol{i};1),\ldots,X^k(\boldsymbol{i};1)),$$

$$(\bar{\eta}_{\boldsymbol{i}}^1,\bar{\eta}_{\boldsymbol{i}}^2,\ldots,\bar{\eta}_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\hat{\boldsymbol{i}};3),X^2(\hat{\boldsymbol{i}};3),\ldots,X^k(\hat{\boldsymbol{i}};3),Y^k(\hat{\boldsymbol{i}};3))$$

$$(\xi_{\boldsymbol{i}}^1,\xi_{\boldsymbol{i}}^2,\ldots,\xi_{\boldsymbol{i}}^{k+1}) \stackrel{\text{def.}}{=} (X^1(\boldsymbol{i};3),X^2(\boldsymbol{i};3),\ldots,X^k(\boldsymbol{i};3),Y^k(\boldsymbol{i};3)),$$

$$(\bar{\xi}_{\boldsymbol{i}}^1,\bar{\xi}_{\boldsymbol{i}}^2,\ldots,\bar{\xi}_{\boldsymbol{i}}^k) \stackrel{\text{def.}}{=} (X^1(\hat{\boldsymbol{i}};1),X^2(\hat{\boldsymbol{i}};1),\ldots,X^k(\hat{\boldsymbol{i}};1)).$$

Then, both $(\eta_{\boldsymbol{i}}^1,\eta_{\boldsymbol{i}}^2,\ldots,\eta_{\boldsymbol{i}}^k)$ and $(\bar{\eta}_{\boldsymbol{i}}^1,\bar{\eta}_{\boldsymbol{i}}^2,\ldots,\bar{\eta}_{\boldsymbol{i}}^{k+1})$ are arithmetic sequences of non-negative integers with common difference -1. Both of the sequences $(\xi_{\boldsymbol{i}}^1,\xi_{\boldsymbol{i}}^2,\ldots,\xi_{\boldsymbol{i}}^{k+1})$ and $(\bar{\xi}_{\boldsymbol{i}}^1,\bar{\xi}_{\boldsymbol{i}}^2,\ldots,\bar{\xi}_{\boldsymbol{i}}^k)$ are arithmetic sequences of non-negative integers with common difference 1.

Then, we can show the required result in a similar way with the proof of Lemma 3. $\square$

The above lemmas directly imply the non-negativity of the transition probability of our joint process.

From the above, we have

$$\mathrm{E}[\ell(X', Y')] = \left(1 - \binom{n}{2}^{-1}\right).$$

It implies the following result.

Theorem 2. *The Markov chain $\mathcal{M}^2$ has the mixing time $\tau_2(\varepsilon)$ satisfying that*

$$\tau_2(\varepsilon) \leqslant (1/2)n(n-1)\ln(dn/(2\varepsilon)),$$

where d is the average of the values in cells, i.e., $d = N/(2^m n)$.

5. Concluding remarks

In this paper, we propose two Markov chains for sampling $(m+1)$-dimensional contingency tables indexed by $\{1, 2\}^m \times \{1, 2, \ldots, n\}$. The first chain has the uniform distribution as a unique stationary distribution. The stationary distributions of the second chain is a conditional multinomial distribution. The mixing times of our chains are bounded by $(1/2)n(n-1)\ln(dn/\varepsilon)$, where d is the average of the values in the cells and ε is a given error bound. Thus, our chains are rapidly mixing. Our result indicates that the mixing times are independent of the dimension $m+1$ of a contingency table when the size is $2\times 2\times\cdots\times 2\times J$.

References

[1] A. Agresti, A survey of exact inference for contingency tables, Statist. Sci. 7 (1992) 131–153.

[2] A. Agresti, Categorical Data Analysis, Wiley, New York, 2002.

[3] D. Aldous, Random walks on finite groups and rapidly mixing Markov chains, In: A. Dold, B. Eckmann (Eds.), Séminaire de Probabilités XVII 1981/1982, Lecture Notes in Mathematics, Vol. 986, Springer, New York, 1983, pp. 243–297.

[4] S. Aoki, Improving path trimming in a network algorithm for Fisher's exact test in two-way contingency tables, J. Statist. Comput. Simulation 72 (2002) 205–216.

[5] S. Aoki, A. Takemura, Minimal basis for connected Markov chain over $3 \times 3 \times K$ contingency tables with fixed two-dimensional marginals, Austral. New Zealand, J. Statist. 45 (2003) 229–249.

[6] R. Bubley, Randomized Algorithms: Approximation, Generation, and Counting, Springer, New York, 2001.

[7] R. Bubley, M. Dyer, Path coupling: a technique for proving rapid mixing in Markov chains, 38th Annu. Symp. on Foundations of Computer Science IEEE, San Alimitos, 1997, pp. 223–231.

[8] F.R.K. Chung, R.L. Graham, S.T. Yau, On sampling with Markov chains, Random Structures and Algorithms 9 (1996) 55–77.

[9] M. Cryan, M. Dyer, L.A. Goldberg, M. Jerrum, R. Martin, Rapidly mixing Markov chains for sampling contingency tables with constant number of rows, Proc. 43rd Annu. Symp. on Foundations of Computer Science (FOCS) 2002, pp. 711–720.

[10] P. Diaconis, B. Efron, Testing for independence in a two-way table: new interpretations of the chi-square statistics (with discussion), Ann. Statist. 13 (1985) 845–913.

[11] P. Diaconis, A. Gangolli, Rectangular arrays with fixed margins, in: D. Aldous, P.P. Varaiya, J. Spencer, J.M. Steele (Eds.), IMA Volumes on Mathematics and its Applications, New York, Springer, Vol. 72, 1995, pp. 15–41.

[12] P. Diaconis, L. Saloff-Coste, Random walk on contingency tables with fixed row and column sums, Technical Report, Department of Mathematics, Harvard University, 1995.

[13] P. Diaconis, B. Strumfels, Algebraic algorithms for sampling from conditional distributions, The Ann. Statist. 26 (1998) 363–397.
[14] M. Dyer, C. Greenhill, Polynomial-time counting and sampling of two-rowed contingency tables, Theoret. Comput. Sci. 246 (2000) 265–278.
[15] M. Dyer, R. Kannan, J. Mount, Sampling contingency tables, Random Structures Algorithms 10 (1997) 487–506.
[16] D. Edwards, T. Havranek, A fast procedure for model search in multidimensional contingency table, Biometrika 72 (1985) 339–351.
[17] R.A. Fisher, Statistical Methods for Research Workers, Oliver and Boyde, Edinburgh, 1934.
[18] R.A. Fisher, The logic of inductive inference (with discussion), J. Roy. Statist. Soc. 98 (1935) 39–54.
[19] D. Hernek, Random generation of $2 \times n$ contingency tables, Random Structures Algorithms 13 (1998) 71–79.
[20] R.W. Irving, M.R. Jerrum, Three dimensional statistical data security problems, SIAM J. Comput. 23 (1994) 170–184.
[21] M.R. Jerrum, A. Sinclair, The Markov chain Monte Carlo method: an approach to approximate counting and integration, in: D.S. Hochbaum (Ed.), Approximation Algorithm for NP-hard Problems, PWS Publishing, Boston, 1997, pp. 482–520.
[22] M. Jerrum, A. Sinclair, E. Vigoda, A polynomial-time approximation algorithm for the permanent of a matrix with non-negative entries, in: Proc. 33rd Annu. ACM Symp. on Theory of Computing, ACM Press, New York, 2001, pp. 712–721
[23] R. Kannan, P. Tetali, S. Vempala, Simple Markov chain algorithm for generating bipartite graphs and tournaments, in: 8th Annu. Symp. on Discrete Algorithms, ACM-SIAM, San Francisco, CA, 1997, pp. 193–200
[24] C.R. Mehta, N.R. Patel, A network algorithm for performing Fisher's exact test $r \times c$ in contingency tables, J. Amer. Statist. Assoc. 78 (1983) 427–434.
[25] K. Pearson, On the χ^2 test of goodness of fit, Biometrika 14 (1922) 186–191.
[26] W.H. Sewell, A.M. Oewnstein, Community of residencial and occupational choice, Amer. J. Sociol. 70 (1965) 551–563.
[27] A. Takemura, S. Aoki, Some characterizations of minimal Markov basis for sampling from discrete conditional distributions, Ann. Inst. Statist. Math. 56 (2004) 1–17.